

Utah Data Research Center Advisory Board

Wednesday December 6th, 2023

Digital Meeting (link on calendar meeting invite)

1:00 pm – 2:00 pm

**Attendees:**

**Board**

Carolyn Dennis

Scott Jones

Rick Little

Carrie Mayne

Chris Williams

**Staff Support**

Anna-Cajsa Gipson

**Attendees**

Zachary Barrus

Andrea Brandley

Jay Chandra

Ari Fenn

Connor Hill

Wei Li

Sarina Ma

Stephen Matherly

Darby McDonough

Alicia McIntire

Jeff Shumway

Jeremias Solari

Karen Tao

Agenda

I. Welcome Jeremias Solari

Meeting Convened at 1:03 pm

II. Review Minutes from Last Meeting Jeremias Solari

Meeting Minutes were sent to Board members and meeting attendees.

**Recommendation:**

No questions or comments; minutes are considered approved.

III. Administrative Update Jeremias Solari

 USBE was awarded an SLDS 2023 Grant for $4 MM; UDRC will be allocated $3 MM of that. The duration of the grant will be 4 years, from October 2023 to September 2027.

The primary objective for the grant is to expand UDRC to handle transactional data moments. Currently, UDRC is optimized for longitudinal research and data management, not transactional data requests.

While the main focus and applications are in the K-20 environment because of the SLDS grant’s educational perspective, we will not be limited to the educational environment.

UDRC will be adding a new product called the Analytics Series to its portfolio of products. In conversations with our data and research partners, we have observed a need for this product.

An analytics product has an optimal delivery time of days to weeks, providing an in-depth analysis on a highly focused topic that fits within the domain of the UDRC. UDRC will allow the latitude for cross agency as well as single agency analysis if the data is internally siloed data and that agency has determined that it is a good use of the UDRC system.

Jeff Shumway would like to clarify the differences between the current process and the new one. Currently, the process is that someone requests a product, then it has to be approved, and then the research done. The new system would be a “rapid fire” sort of process that is much shorter than the current one. In the new system, a request would be prioritized, and completed within days or weeks.

The type of product would involve an operational question or a question that is temporally sensitive.

This should solve a demand for research by our data partners. One of the areas of need is for legislative requests to be processed quickly, which could occur in the new system.

Darby would like to add that Commerce is working with DTS to get pipelines up so that they can get a data sharing agreement with other stakeholders.

Chris is wondering if the requests are limited to data partners or if anyone can reach out to UDRC. He also wants to clarify that data partners will be informed that their data will be used in potential products.

This will add value to what UDRC is currently producing.

UDRC’s governance document was last updated in 2020 and is in need of an update. When UDRC migrated over to USHE, there were funds allocated for a data privacy position, which has been filled by Jeff Johnson at USHE. Jeff has helped update the verbiage surrounding data privacy in the document.

Executive Order 2023-01 created a task force to look into state agency collaboration as it relates to data sharing. UDRC had an informal role by being part of two sub-committees.

UDRC contributed expertise relating to data sharing: one perspective was regarding enterprise data where all data belongs to the executive branch and therefore its utilization should be governed within an enterprise data framework. The other was in the centralization of efforts to minimize duplication and therefore cost to the state.

UDRC will be hosting a Bagels & Research presentation in the first quarter of 2024. UDRC will provide updates to the research community on current and upcoming research.

**Recommendations:**

Allocation of $3 MM will be towards:

* Third Party IT Services
* Personnel expansion
* Incidentals for grant-related activities

Example use cases:

* SSID Services across institutions
* Transcript Services
* Common Admissions Application to make determinations for student applications

Jeremias will attend the grant orientation meeting in Washington, D.C. in December.

UDRC to review the results of the legislative session to confirm potential progress on receiving licensure data

UDRC has completed a framework surrounding the process for requesting a longitudinal analytics product:

* Form
* SOP
* Templates
* Website components

Full implementation of this product line will occur in the next month.

Data requests can be made by anyone who reaches out to UDRC. UDRC will pre-screen projects before starting them. Data partners will be informed when their data will be used for a request, and they can voice questions, comments, and concerns.

The UDRC governance document will reflect UDRC’s move from DWS to USHE and the addition of the Commerce Department as a data partner. The document will be more specific around cybersecurity and data privacy. UDRC is finalizing the first draft. The document will then go to peer review, finalization, and publication on the UDRC website.

The task force, that UDRC was not formally part of, will provide recommendations to the legislative and executive branches.

UDRC will reach out to data and research partners for the possibility of a shared presentation.

IV. Technology Update Jeremias Solari

UDRC’s phase V system’s formal development work was completed in September 2023 and moved to user acceptance testing in October 2023. User acceptance testing was completed for the UDRC core, which comprises an export application, ingest pipeline, matching with synthetic data sets, and de-identification.

The UDRC core will be deployed into a production environment where it is possible to receive data from data partners. Prod Zero was completed this week.

Chris asked about the timeframe UDRC is targeting to receive the data set from data partners.

Chris wants to know who provides maintenance when data partners transfer data and it is stored in the UDRC system.

With the UDRC system holding data elements, Chris is concerned who has access to the raw data provided by partners. DWS is very interested in knowing who has access to data because of how the MOU is currently written. The current MOU does not specify users, so they would like to coordinate who has access to that data for an expectation of security.

Raw data access is very well established and our contracts with Burwood have been reviewed by UDRC’s data privacy officer. UDRC has already implemented structure for data access.

A new implementation of the export application was put in place. It is the official bespoke data transfer tool and is a cloud compliant version. Additional security has been added. The importance of the export application is that it will authenticate the data sender, which allows UDRC to specify who can send data to the system.

UDRC will implement process automation and have full control of the code. UDRC has full control of security and code, rather than having to use products like MoveIt.

DWS has also submitted for data security and DTS review for the phase V system.

The MPI algorithm has gone through a shift in perspective from phase IV to phase V. In phase IV, it was deterministic with a simplistic matching function. In phase V, the system moved to a hybrid deterministic/probabilistic approach. UDRC is using a classifying function (gradient boosting classifier) that allows flexibility for noisy data and allows assigning disparate weighting of data variables like SSN.

UDRC has changed the way that name comparison works to be more flexible instead of binary. The approach now is to use a probabilistic perspective using Levenshtein distance to calculate how similar one name is to another name.

A new search function was also implemented. In Phase IV, UDRC used SQL stored procedures with native search functions. Phase V uses a vectorized search tree that shows only those MPIs in the system that have a high level of similarity, which decreases search time.

UDRC has aligned with a framework for cybersecurity called FedRAMP. We have targeted FedRAMP Moderate, which is applicable for when there would be a loss of confidentiality, integrity, and availability that would result in serious adverse effects on an agency’s operations, assets, or individuals. For context, FedRAMP high is reserved for law enforcement, emergency services, health, and financial systems.

New features in the phase V system are things like:

* Modern analytics workbench that is highly connected to the data, scalable, and has modern tooling
* Data partner access is enhanced so that the data partner can specify individuals that can access their own data

Instead of requesting a yearly file, data partners could request data from other data partners on their own within the UDRC system.

Bridge expansion is four years starting in October of this year.

**Recommendations:**

UDRC’s phase V system is currently working on warehouse operations, user interface, and external data request process.

Next steps are to complete regression testing. At that point, UDRC will be ready to accept and process raw data from partners. Researchers will be able to have the most up to date data and provide data to individuals who request it.

individuals who request it.

UDRC will reach out to data partners in the next week or two in regard to providing data sets for system testing.

UDRC data engineers will provide maintenance for the system, but Burwood, a 3rd party vendor will provide assistance for deeper level system features.

UDRC data engineers, Jeremias, and the Burwood group will have their ability to access raw data removed once the active development process is complete. Per UDRC’s governance document.

User acceptance will not be concurrent with data partners submitting raw data. UDRC will create a process where data partners would be notified if Burwood needed to access the system for updates that would include raw data.

Develop a contingency plan for timeframes of if data partners cannot submit data on a case-by-case basis.

UDRC is working on integrating batch processing for efficiency and speed.

Commerce, USBE, USHE, and DHHS have all installed and tested the system. UDRC is waiting for DTS review by DHHS, and it should be completed in the next month or two.

Those who have tested the product have provided feedback to UDRC and it has been implemented. Jeremias will reach out to Darby to set up a meeting to provide documentation on how to replicate the testing. Commerce did install and test the system under Darby’s predecessor. The connection was tested. Jeremias will still connect with Darby. Rick will ask Valli to share with Chris some of the feedback from their testing.

A nuanced primer developed by UDRC on the MPI will be publicly available by the end of next week. The results of the synthetic data set testing are available now.

Align with FedRAMP moderate, which will provide guidance for how the UDRC system should handle and look foundationally and structurally and what sort of features the system should have on an infrastructure, system, and operational level code for the use of the system.

Discuss the possibility of implementing a feature where data partners will have access to other data partners’ data.

Create a process for how this would work and what specific purposes data partners can have access.

The Bridge expansion timeline is as follows:

* Year 1 will focus on tangible requirements gathering and personnel onboarding
* Years 2 & 3 will begin formal development work and delivery of the Bridge system into user acceptance training
* Year 4 should see the completion of user acceptance training, onboarding of data requesters and providers and the system should exist in a production environment

V. Research Update

Personnel update: Alicia McIntire has joined UDRC in September as a researcher.

 V.1 – Stacked Awardees Karen Tao

This research has been completed and was published on the UDRC website on August 16th.

**Recommendations:**

For further information on this research, please visit the UDRC website.

Gender Wage Gap & WOC Karen Tao

Currently, we are waiting for the USHE comms team to review the report. Peer review and data narrative have both been completed.

**Recommendations:**

Karen will conduct final internal reviews by incorporating feedback and then launch on the UDRC website.

 Charter Schools Karen Tao

Karen met with data partners for initial meetings. The SQL code for the data pull has been completed and definitions have been aligned.

 **Recommendations:**

Karen will continue with the analysis and statistical modeling. She will then go over any findings with data partners and make any adjustments of the analysis and modeling code. After that, the first draft of the narrative will be generated for more peer review.

V.2 – Areas of Study, COVID, & Wages Outcomes Connor Hill

Connor is looking at how the pandemic impacted different demographic groups as well as different outcomes for degree or certificate recipients. Peer review feedback has been implemented and final report completed.

**Recommendations:**

Connor will go through final editing steps before the report is published.

 Cost of Living Connor Hill

This research focuses on the number of USHE graduates in Utah who are able to afford basic necessities. Connor has met with data partners and has received the data. He is in the final steps of the analysis.

 **Recommendations:**

Connor will meet with data partners again to discuss the analysis. The report will then go through internal peer review and a first draft will be generated.

V.3 – Prior Learning Assessment Alicia McIntire

Alicia introduced herself as a new researcher for UDRC. This research will focus on AP and CLEP as that is where the most data is. Data collection is finished, and analysis is moving forward.

**Recommendations:**

Alicia is meeting with data partners to move forward with the analysis.

V.2 – Firm Size and Wage Growth Ari Fenn

There is a wage premium for those who are employed at larger firms. This research has been completed and was published on the UDRC website on November 28th.

**Recommendations:**

For further information on this research, please visit the UDRC website.

Nursing Occupational Pathways Ari Fenn

This research report has been drafted and is currently in internal review. Carrie indicated that while Ari did not coordinate directly with HWAC and HWIC, UDRC has representation on both of those levels and indirectly coordinated with them. Jeremias indicated that in terms of the interim presentations, both groups were invited to those secondary meetings.

**Recommendations:**

Following internal review, Ari will conduct data partner review before sending the report to UX to create a data narrative. UDRC may take the results of this study to both HWAC and HWIC. Sync up with UDRC, DHHS, and HWAC per Jeff’s suggestion.

VI. UX Research Update Zachary Barrus

Work done by the UX team is mostly related to the UDRC website. Since the last meeting, there have been several updates to the website:

* Data narratives for completed research
* Improvements to the UDRC website to make it more user friendly and accessible. Accessibility features have been standardized. Aesthetics have also been standardized
* Improved user experience by listing research in reverse chronological order and linking data narratives to the research agenda

**Recommendations:**

Create and publish data narratives for the UDRC analytics series.

Develop a UX policy that sets out expectations for products:

* Accessibility
* Data security
* Interactivity
* Equity in representing different groups
* Aesthetics
* Formatting

VII. Data Request Update Jeremias Solari

The data request from Salt Lake County has been completed. This request’s purpose was to approximate income for WISE participants before and after attaining certificates from the University of Utah and Salt Lake Community College for specific CIP codes. That data was communicated to Salt Lake County on November 7th and the data request is now complete.

**Recommendations:**

Salt Lake County can now go into the product development phase and then to a peer review process.

VIII. Public Comment

Meeting adjourned at 2:01 pm